

Software

Open Access

Storing, linking, and mining microarray databases using SRS

Antoine Veldhoven¹, Don de Lange¹, Marcel Smid², Victor de Jager³, Jan A Kors⁴ and Guido Jenster*¹

Address: ¹Department of Urology, Josephine Nefkens Institute, Erasmus MC, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, ²Medical Oncology, Erasmus MC, Rotterdam, The Netherlands, ³Bioinformatics, Erasmus MC, Rotterdam, The Netherlands and ⁴Medical Informatics, Erasmus MC, Rotterdam, The Netherlands

Email: Antoine Veldhoven - antoine@b2bdsl.nl; Don de Lange - d.delange@erasmusmc.nl; Marcel Smid - m.smid@erasmusmc.nl; Victor de Jager - victor.de.jager@nbic.nl; Jan A Kors - j.kors@erasmusmc.nl; Guido Jenster* - g.jenster@erasmusmc.nl

* Corresponding author

Published: 27 July 2005

Received: 06 May 2005

BMC Bioinformatics 2005, 6:192 doi:10.1186/1471-2105-6-192

Accepted: 27 July 2005

This article is available from: <http://www.biomedcentral.com/1471-2105/6/192>

© 2005 Veldhoven et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: SRS (Sequence Retrieval System) has proven to be a valuable platform for storing, linking, and querying biological databases. Due to the availability of a broad range of different scientific databases in SRS, it has become a useful platform to incorporate and mine microarray data to facilitate the analyses of biological questions and non-hypothesis driven quests. Here we report various solutions and tools for integrating and mining annotated expression data in SRS.

Results: We devised an Auto-Upload Tool by which microarray data can be automatically imported into SRS. The dataset can be linked to other databases and user access can be set. The linkage comprehensiveness of microarray platforms to other platforms and biological databases was examined in a network of scientific databases. The stored microarray data can also be made accessible to external programs for further processing. For example, we built an interface to a program called Venn Mapper, which collects its microarray data from SRS, processes the data by creating Venn diagrams, and saves the data for interpretation.

Conclusion: SRS is a useful database system to store, link and query various scientific datasets, including microarray data. The user-friendly Auto-Upload Tool makes SRS accessible to biologists for linking and mining user-owned databases.

Background

The extraction of information from data generated by high-throughput experiments in genomics and proteomics has been likened to "attempting to drink from a fire hose". We are flooded with information on many levels such as whole genome DNA sequences, RNA expression, protein-protein interactions, protein modifications, and more. All this information is accessible in very different formats, ranging from well-organized curated gene sequences to unstructured free text in scientific literature.

A system that can manage, link and query these heterogeneous types of datasets is therefore extremely valuable. The Sequence Retrieval System (SRS) is such a unified database system in which numerous different scientific databases have already been integrated [1].

Of special interest are data from high-throughput RNA expression microarrays [2,3]. Many of these datasets are freely available and, like information stored in other scientific databases, are from different platforms [4,5].

Integrating and mining these databases strongly facilitates the analysis of genes of interest but will also support discovery of disease markers, drug-targets and new knowledge in general [6-9]. One such platform is Oncomine, which has integrated many different microarray datasets, focussing on human cancer [10]. Additionally, standardized microarray depositories such as GEO (Gene Expression Omnibus) [11], ArrayExpress [12], and CIBEX [13] do or will soon provide options to browse and query the datasets [14-17]. No doubt, other platforms will be developed focussing on the integration of microarray data. If started from scratch, these initiatives will likely be limited in their direct linkage to other heterogeneous biological databases due to the laborious task of making those connections and programming the single and batch-wise query options. The universality and the availability of numerous scientific databases that have already been integrated in SRS make it a useful platform for integrating microarray databases. Although the SRS interface to query databases is quite user friendly, other aspects of working with SRS are not. These include (i) uploading microarray datasets, (ii) database security including setting user access, (iii) linking databases, (iv) generating standard views, and (v) communication with other programs such as statistical and clustering software. The current SRS interface has a major disadvantage in that it is not designed to perform complex calculations on the fly. This means that any microarray dataset to be uploaded must have all ratio and statistical calculations performed upfront. For example, once in SRS, one cannot change ratios from log10 to log2 or add an extra field per gene by dividing expression data of all "normal" by all "cancer" samples. However, software programs that perform calculations, statistical evaluations, clustering, protein domain predictions, homology searches, and more, can communicate with SRS. Interfaces can be generated that retrieve data from SRS, perform the required action and if desired, store the results in SRS. Alternatively, SRS allows direct integration of programs such as the BLAST and FASTA homology searches and the SRS-EMBOSS (European Molecular Biology Open Software Suite) tools [18,19].

Generating a database in which heterogeneous datasets are integrated is a challenge in itself. However, retrieving statistically meaningful data by comparing datasets from different sources, platforms and designs is particularly difficult [20]. There is a fast growing body of publications on microarray cross-platform comparisons, mainly showing how this can be achieved in very many different ways [8,21-26]. Statistical evaluations of data within a dataset of sufficient technical and biological replicates, are better defined and can be implemented per dataset within a database system [27,28]. The strategies and applications we discuss here to link, store and query scientific datasets in SRS, do not go beyond processed individual datasets

and do not include cross-platform dataset integrations. We assume that each uploaded dataset consists of high-quality data and has been processed correctly.

In this paper we describe strategies to incorporate microarray databases into SRS and provide a database upload tool. Using the program Venn Mapper as an example, we show the possibility to automatically retrieve the stored microarray data from SRS for external statistical evaluation.

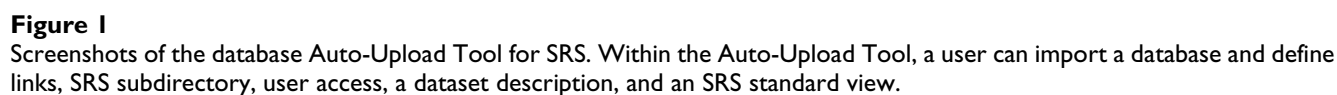
Implementation

Auto-Upload Tool

In order to import microarray databases into SRS (version 7.1.3), an Auto-Upload Tool was built (Figure 1). This PHP-written tool allows one to store databases of a predefined format into a user-owned and password protected directory on a local SRS server [see Additional file 1] [29]. In this directory, databases can be managed, viewed and uploaded into SRS. In the "edit-database" interface, links to other databases can be specified for each field. A standard view can be generated and the location of the dataset in the SRS directory determined. Finally, permissions can be set to control access to the various datasets in SRS. Upon uploading, the Auto-Upload Tool will generate the files required for SRS: (i) the SRS data file in which the spreadsheet input file is converted into a flat-file database, (ii) an Icarus syntax file (.is-file) which describes the layout of the flat-file database, (iii) a database index file (.i-file) which describes the way in which the different fields need to be indexed for the SRS system, (iv) a database view file (.view-file) in which a standard view is defined, and (v) an information file (.it-file) which can harbour a description of the dataset. These files are automatically placed into the SRS directory after which the Auto-Upload Tool updates the *rsdb.i*, *user.i* and *site.i* files. These files describe the name of the database, where the files are located (*rsdb.i*), user permissions (*user.i*), and configuration of the different database groups (*site.i*). The *srsc* command within the Auto-Upload Tool implements the changes in the configuration files after which *srsc* and *rsdo* perform indexing of new databases and set links. Incorporation of a new dataset using this tool generally takes place within minutes. On our local SRS server, a DQS (Distributed Queuing System) batch-queue is installed to prevent data loss or corruption of datasets in case multiple users are editing datasets at the same time.

External programs accessing SRS: Venn Mapper for SRS

An important feature of storing and linking microarray data in SRS is the accessibility of the datasets for other programs. As an example, we generated a PHP web interface for the Venn Mapper program that retrieves microarray data from SRS to calculate the statistical significance of the number of co-occurring differentially expressed genes in



bling the use of different ratio cut-offs for different microarray experiments. Upon login, the interface dis-

plays all microarray databases indexed in SRS to which the user has access. After selection of the datasets a second screen shows all fields (such as individual array experiments or averaged group ratios) of the selected databases. The microarray experiments of interest can be selected for Venn Mapper analysis after which the requested data is linked and exported from SRS into the Venn Mapper program. The output of the program is available for viewing and downloading. Information requests from the interface to SRS are made through the SRS getz command [30]. This powerful feature of SRS makes the integrated databases accessible to any external program.

Results and discussion

Preparing and linking of microarray databases in SRS

With respect to the microarray database set-up, there are two important considerations. First, in our experience, microarray data mining often starts with selecting genes based on their differential expression. Differential gene expression is best determined using statistical evaluation of the data based on sufficient technical and biological replicates [27,28]. Dependent on the statistical test and microarray platform, raw gene expression data and/or ratio calculations can be utilized. Second, making changes to a microarray dataset in SRS is impractical and datasets should be fully built before they are imported. This means that raw expression data and ratio calculations should be normalised and flagged and represented in a common format (such as log2). Importantly, statistical evaluation should be included. For simplicity in representation, datasets can be summarised in, for example, an average of all "normal" and "cancer" samples and additional fields of log2 "normal/cancer" can be included.

Linking of databases should be based on invariable and unique indexes. Links based on identifiers such as UniGene cluster identifiers that are regularly re-assigned, forces one to repeatedly update all databases that include such a denominator. Invariable links based on DNA sequence assignments such as GenBank and ProbeSet identifiers (IDs) are therefore more appropriate linking indexes. We recommend including only one of those invariable and unique linking fields in the microarray dataset to avoid the need for a regular update. Since many biological databases do not use these hard links, connecting microarray datasets to other databases can be achieved through a coupling file and/or by making use of the web of links provided by the biological databases (Figure 2). A coupling file can be minimal, containing for example only a RefSeq ID with the appropriate array spot number, or can contain a variety of indexes such as GenBank, RefSeq, SwissProt, OMIM, LocusLink, KEGG, GO, GeneCards, UniGene identifiers, that directly link to the different databases. Since some of these links are variable and biological databases keep growing with information, cou-

pling files must be regularly renewed to update the various links.

Instead of or in combination with a coupling file, one can make use of the links provided in the various biological databases (Figure 2). For example, the GenBank accession code in the microarray dataset can be linked to the UniGene database. The LocusLink database is linked to the UniGene database through RefSeq accession numbers and also contains IDs that for example link to EMBL, OMIM, and SwissProt. In this way, almost all biological databases are linked through a network of direct and indirect connections. In case multiple roads lead to the same databases, SRS utilizes only one route. By assigning values to each link, the route taken is the one having the lowest sum of link values, even when this results in a lower number of connected fields (genes). Although SRS can be forced to take a specified path, one should be careful to be dependent on many different databases. Inconsistencies in and incompleteness of databases are accumulated when linking occurs in sequence.

Linking to the network of biological databases through a coupling file has various advantages. One can establish directly validated links to each database, including databases outside the network. In addition, errors in links can easily be corrected. Platform-specific or overall coupling files can be retrieved from the Affymetrix website [31] and from sources such as Resourcer [32], KARMA [33], GeneHopper [34], ProbeMatchDB [35], DAVID [36], Ensembl [37], and Source [38]. Using these resources, microarray datasets from different platforms can be linked. This includes connecting cross-species datasets using ortholog converters such as HOMGL [39] and HOMOLOGENE [40].

Gene linking efficiency of different databases

A high accuracy and comprehensiveness of linking are essential for a successful comparison of microarray data from different platforms. The extent of linkage of various databases in SRS was examined (Figure 3). The percentage of fields of scientific databases and microarray datasets that are linked to other databases was assessed. As shown in Figure 2, most microarray databases are directly linked to the scientific database network via a single connection. The U133A and U95 Affymetrix coupling files contain direct links to UniGene, SwissProt, LocusLink, and OMIM. The linkage of the various microarray platforms to UniGene varies between 84% and 96%. On average, 60% of the genes of microarray datasets can be linked to each other via UniGene.

Auto-Upload Tool and external programs accessing SRS

The availability of many scientific databases in SRS, the universality of the system and its free access for academic

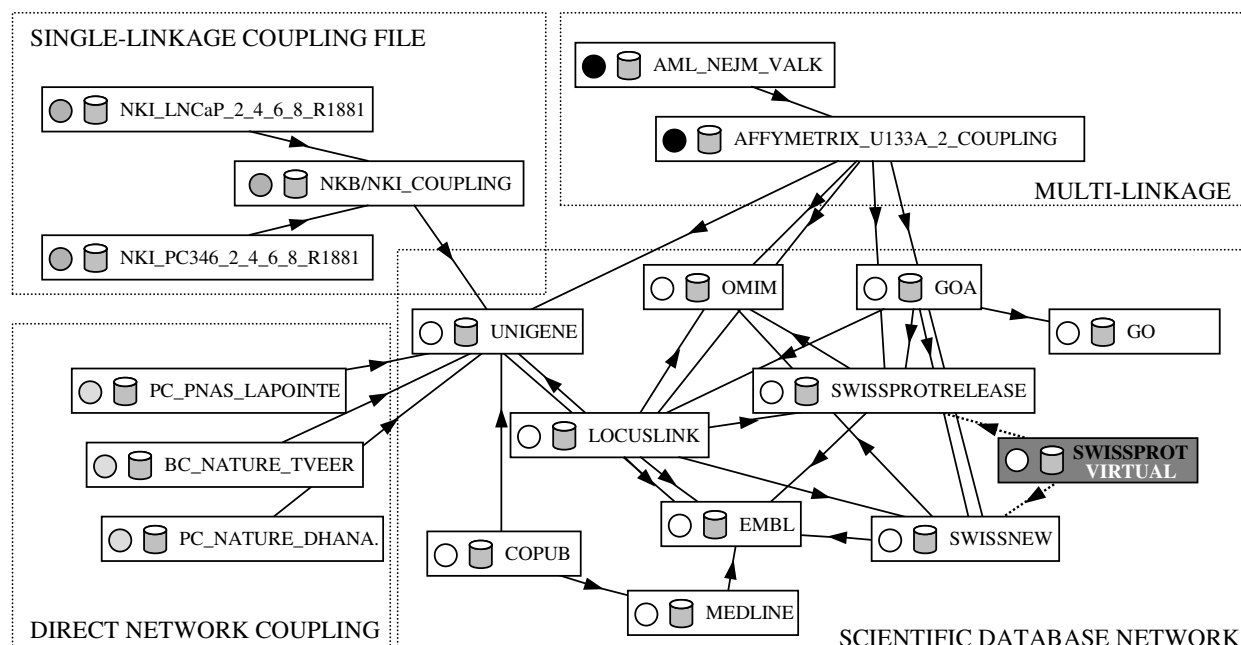


Figure 2

SRS Universe example view of database linkage. The scientific database network consists of well-known interlinked biological databases. Microarray datasets are coupled to this network through a multi-linked Affymetrix coupling file, the single UniGene-linked NKB/NKI-coupling file, or directly to UniGene using GenBank accession numbers. For clarity, not all databases and links mentioned are represented in the scheme. Public SRS servers and indexed scientific databases can be viewed at [43].

use, make SRS an excellent mining system for heterogeneous microarray datasets. The Auto-Upload Tool facilitates the exchange of microarray datasets between separate SRS installations. Using a single data file and optional description file, any user can upload the identical data and customize it to their own SRS environment. We would urge researchers and microarray data repositories to make their data available in an SRS format. In addition, microarray software programmers could make their software available in an SRS compatible format or include SRS data export options. The commercial SRS GeneSpring® Connector and public EMBOSS are examples of such microarray-SRS integration ventures.

We plan to extend our efforts of integrating more microarray databases into SRS. In addition, software tools specific for microarray data analysis, such as Go Mapper and CoPub Mapper will be rewritten for SRS [26,41,42]. The CoPub Mapper literature mining program contains databases that store, for each gene, all MEDLINE records mentioning the gene. This directly links microarray expression

data to the published literature and allows for co-publication research of gene-gene and gene-keyword combinations.

Conclusion

The Sequence Retrieval System is a versatile and useful database system to store, link and query various scientific databases, including microarray datasets. Fully processed datasets can be incorporated and linked to other datasets using the Auto-Upload Tool. This user-friendly program makes SRS accessible to users who can themselves add, link and mine databases within minutes. Datasets stored in SRS can be interrogated by external programs to perform virtually any computation.

Availability and requirements

Project Name: Auto-Upload Tool and Venn Mapper for SRS

Project home page: <http://www.erasmusmc.nl/gatcplatform>

<div>LINKING</div> <div>TO: →</div> <div>FROM: ↓</div>	UNIGENE (human) (107,014)	SWISSPROT (human) (11,164)	ENSEMBL (human) (3,459)	LOCUSLINK (human) (38,430)	OMIM (all records including diseases) (15,915)	Gene Ontology (13,238)	Affymetrix U133A 2.0 Coupling (22,282)	Affymetrix U95 Av2 Coupling (12,626)	NKB/NKI Coupling (18,758)	BC_NATURE_tVeer (24,522)	Incyte UNIGEM V (9,698)	PC_NATURE_Dhana. (12,678)	Sigma/Compugen oligo array (18,660)	PC_PNAS_Lapointe (20,165)
UNIGENE (human, build 167) (107,014)		10%	16%	19%	9%	12%	12%	8%	12%	18%	7%	6%	14%	11%
SWISSPROT (human, release 42) (11,164)	92%		96%	94%	96%	86%	77%	62%	63%	82%	49%	36%	75%	52%
ENSEMBL (human, build 34) (3,459)	82%	72%		83%	75%	77%	76%	71%	73%	78%	67%	61%	77%	70%
LOCUSLINK (human, March 2004) (38,430)	54%	27%	48%		28%	35%	33%	23%	29%	39%	20%	15%	34%	25%
OMIM (March 2004) (15,915)	59%	77%	59%	68%		55%	53%	43%	41%	54%	33%	24%	50%	34%
Gene Ontology (March 2004) (13,238)	100%	73%	94%	100%	68%		64%	53%	62%	83%	47%	34%	75%	53%
Affymetrix U133A 2.0 Coupling (22,282)	96%	65%	92%	95%	65%	71%		73%	69%	84%	53%	39%	80%	60%
Affymetrix U95 Av2 Coupling (12,626)	95%	73%	92%	95%	75%	74%	74%		75%	88%	62%	44%	82%	61%
NKB/NKI Coupling (18,758)	93%	52%	76%	79%	49%	62%	66%	54%		76%	47%	47%	66%	68%
BC_NATURE_tVeer (24,522)	93%	46%	70%	74%	44%	57%	59%	43%	54%		38%	30%	64%	49%
Incyte UNIGEM V (9,698)	89%	62%	81%	83%	60%	71%	75%	67%	69%	81%		41%	73%	61%
PC_NATURE_Dhana. (12,678)	84%	55%	69%	72%	54%	62%	64%	57%	74%	71%	49%		62%	62%
Sigma/Compugen oligo array (18,660)	89%	51%	73%	76%	49%	61%	66%	48%	53%	74%	39%	29%		47%
PC_PNAS_Lapointe (20,165)	84%	45%	70%	73%	42%	55%	59%	45%	68%	70%	42%	41%	60%	

Figure 3

Linkage efficiency of databases within SRS. The percentage of records in a specific database (in rows) linked to others (columns) are shown. Numbers in brackets are the total number of fields in the particular database. The NKB/NKI Coupling [44], BC_NATURE_tVeer [45], Incyte Human UNIGEM V 1.0, PC_NATURE_Dhana. [46], Sigma/Compugen oligo array [47], and PC_PNAS_Lapointe [48], are linked to UniGene via a single accession code connection (see Figure 2). Direct links are depicted as grey cells. The Affymetrix U133A and U95 coupling files [31] are directly linked to different biological databases.

Operating system: Platform independent

Programming language: PHP, JavaScript, Perl

Other requirements: Local SRS installation, DQS batch-queue, MySQL database server, PHP-enabled Webserver (like Apache)

License: SRS (Lion Bioscience)

Any restrictions to use by non academics: License needed

Authors' contributions

AV and DdL generated the Auto-Upload Tool. DdL, AV and MS generated the Venn Mapper for SRS program. AV and VdJ installed and managed the servers for the various tools. Funding for the project was obtained by JK and GJ. AV, JK and GJ contributed to the intellectual content and GJ supervised the project.

Additional material

Additional File 1

describing how to use the Auto-Upload Tool program

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-6-192-S1.pdf>]

Acknowledgements

We would like to thank EMBL/EBI and Lion Bioscience for making SRS available and Peter Hendriksen for careful reading of the manuscript. This work was supported by Erasmus MC Breedtestrategie and the Urologic Research Foundation (SUWO) Erasmus MC.

References

- Zdobnov EM, Lopez R, Apweiler R, Etzold T: **The EBI SRS server – recent developments.** *Bioinformatics* 2002, **18**:368-373.
- Brown PO, Botstein D: **Exploring the new world of the genome with DNA microarrays.** *Nat Genet* 1999, **21**:33-37.
- Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM: **Expression profiling using cDNA microarrays.** *Nat Genet* 1999, **21**:10-14.
- Heller MJ: **DNA microarray technology: devices, systems, and applications.** *Annu Rev Biomed Eng* 2002, **4**:129-153.
- Schena M, Heller RA, Thieriault TP, Konrad K, Lachenmeier E, Davis RW: **Microarrays: biotechnology's discovery platform for functional genomics.** *Trends Biotechnol* 1998, **16**:301-306.
- Moreau Y, Aerts S, De Moor B, De Strooper B, Dabrowski M: **Comparison and meta-analysis of microarray data: from the bench to the computer desk.** *Trends Genet* 2003, **19**:570-577.
- Rhodes DR, Chinnaiyan AM: **Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers.** *Ann N Y Acad Sci* 2004, **1020**:32-40.
- Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: **Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression.** *Proc Natl Acad Sci U S A* 2004, **101**:9309-9314.
- Welsh JB, Sapinoso LM, Kern SG, Brown DA, Liu T, Bauskin AR, Ward RL, Hawkins NJ, Quinn DI, Russell PJ, Sutherland RL, Breit SN, Moskaluk CA, Frierson HF Jr, Hampton GM: **Large-scale delineation of secreted protein biomarkers overexpressed in cancer tissue and serum.** *Proc Natl Acad Sci U S A* 2003, **100**:3410-3415.
- Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: **ONCOMINE: a cancer microarray database and integrated data-mining platform.** *Neoplasia* 2004, **6**:1-6.
- Edgar R, Domrachev M, Lash AE: **Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.** *Nucleic Acids Res* 2002, **30**:207-210.
- Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA: **ArrayExpress – a public repository for microarray gene expression data at the EBI.** *Nucleic Acids Res* 2003, **31**:68-71.
- Ikeo K, Ishi-i J, Tamura T, Gojobori T, Tateno Y: **CIBEX: center for information biology gene expression database.** *C R Biol* 2003, **326**:1079-1082.
- Stoeckert CJ Jr, Causton HC, Ball CA: **Microarray databases: standards and ontologies.** *Nat Genet* 2002, **32**:469-473.
- Gardiner-Garden M, Littlejohn TG: **A comparison of microarray databases.** *Brief Bioinform* 2001, **2**:143-158.
- Quackenbush J: **Data standards for 'omic' science.** *Nat Biotechnol* 2004, **22**:613-614.
- Penkett CJ, Bahler J: **Getting the most from public microarray data.** *European Pharmaceutical Review* 2004, **9**:8-17.
- Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**:276-277.
- Kulikova T, Aldebert P, Althorpe N, Baker W, Bates K, Browne P, van den BA, Cochrane G, Duggan K, Eberhardt R, Faruque N, Garcia-Pastor M, Harte N, Kanz C, Leinonen R, Lin Q, Lombard V, Lopez R, Mancuso R, McHale M, Nardone F, Silventoinen V, Stoeck P, Stoesser G, Tuli MA, Tzouvara K, Vaughan R, Wu D, Zhu W, Apweiler R: **The EMBL Nucleotide Sequence Database.** *Nucleic Acids Res* 2004, **32**:D27-D30.
- Marshall E: **Getting the noise out of gene arrays.** *Science* 2004, **306**:630-631.
- Zhou XJ, Kao MC, Huang H, Wong A, Nunez-Iglesias J, Primig M, Aparicio OM, Finch CE, Morgan TE, Wong WH: **Functional annotation and network reconstruction through cross-platform integration of microarray data.** *Nat Biotechnol* 2005, **23**:238-243.
- Mitchell SA, Brown KM, Henry MM, Mintz M, Catchpoole D, LaFleur B, Stephan DA: **Inter-platform comparability of microarrays in acute lymphoblastic leukemia.** *BMC Genomics* 2004, **5**:71.
- Chiorino G, Acquadro F, Mello GM, Viscomi S, Segir R, Gasparini M, Dotto P: **Interpretation of expression-profiling results obtained from different platforms and tissue sources: examples using prostate cancer data.** *Eur J Cancer* 2004, **40**:2592-2603.
- Culhane AC, Perriere G, Higgins DG: **Cross-platform comparison and visualisation of gene expression data using co-inertia analysis.** *BMC Bioinformatics* 2003, **4**:59.
- Shippy R, Sendera TJ, Lockner R, Palaniappan C, Kayser-Kranich T, Watts G, Alsobrook J: **Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations.** *BMC Genomics* 2004, **5**:61.
- Smid M, Dorssers LC, Jenster G: **Venn Mapping: clustering of heterogeneous microarray data based on the number of co-occurring differentially expressed genes.** *Bioinformatics* 2003, **19**:2065-2071.
- Cui X, Churchill GA: **Statistical tests for differential expression in cDNA microarray experiments.** *Genome Biol* 2003, **4**:210.
- Draghici S: **Statistical intelligence: effective analysis of high-density microarray data.** *Drug Discov Today* 2002, **7**:S55-S63.
- Auto-Upload Tool Manual [<http://www.erasmusmc.nl/gatcplatform/autouploadmanual.pdf>]
- Schaftenaar G, Cuelenaere K, Noordik JH, Etzold T: **A Tcl-based SRS v. 4 interface.** *Comput Appl Biosci* 1996, **12**:151-155.
- Affymetrix [<http://www.affymetrix.com>]
- Tsai J, Sultana R, Lee Y, Perlea G, Karamycheva S, Antonescu V, Cho J, Parvizi B, Cheung F, Quackenbush J: **RESOURCEER: a database for annotating and linking microarray resources within and across species.** *Genome Biology* 2001, **2**:software0002.
- Cheung KH, Hager J, Pan D, Srivastava R, Mane S, Li Y, Miller P, Williams KR: **KARMA: a web server application for comparing and annotating heterogeneous microarray platforms.** *Nucleic Acids Res* 2004, **32**:V441-V444.
- Svensson BA, Kreeft AJ, van Ommen GJ, den Dunnen JT, Boer JM: **GeneHopper: a web-based search engine to link gene-expression platforms through GenBank accession numbers.** *Genome Biol* 2003, **4**:R35.
- Wang P, Ding F, Chiang H, Thompson RC, Watson SJ, Meng F: **ProbeMatchDB – a web database for finding equivalent probes across microarray platforms and species.** *Bioinformatics* 2002, **18**:488-489.
- Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4**:P3.

37. Kasprzyk A, Keefe D, Smedley D, London D, Spooner W, Melsopp C, Hammond M, Rocca-Serra P, Cox T, Birney E: **EnsMart: a generic system for fast and flexible access to biological data.** *Genome Res* 2004, **14**:160-169.
38. Diehn M, Sherlock G, Binkley G, Jin H, Matese JC, Hernandez-Bousard T, Rees CA, Cherry JM, Botstein D, Brown PO, Alizadeh AA: **SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data.** *Nucleic Acids Res* 2003, **31**:219-223.
39. Bluthgen N, Kielbasa SM, Cajavec B, Herzel H: **HOMGL-comparing genelists across species and with different accession numbers.** *Bioinformatics* 2004, **20**:125-126.
40. Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Suzek TO, Tatusova TA, Wagner L: **Database resources of the National Center for Biotechnology Information: update.** *Nucleic Acids Res* 2004, **32**:D35-D40.
41. Alako BT, Veldhoven A, van Baal S, Jelier R, Verhoeven S, Rullmann T, Polman J, Jenster G: **CoPub Mapper: mining MEDLINE based on search term co-publication.** *BMC Bioinformatics* 2005, **6**:51.
42. Smid M, Dorssers LC: **GO-Mapper: functional analysis of gene expression data using the expression level as a score to evaluate Gene Ontology terms.** *Bioinformatics* 2004, **20**:2618-2625.
43. **Public SRS servers** [http://downloads.lionbio.co.uk/public_srs.html]
44. **NKI Central Microarray Facility** [<http://microarrays.nki.nl/>]
45. 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der KK, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**:530-536.
46. Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurchi K, Pienta KJ, Rubin MA, Chinnaiyan AM: **Delineation of prognostic biomarkers in prostate cancer.** *Nature* 2001, **412**:822-826.
47. **Compugen Oligo Library** [<http://www.labonweb.com/chips/libraries.html>]
48. Lapointe J, Li C, Higgins JP, Van de RM, Bair E, Montgomery K, Ferrari M, Egevad L, Rayford W, Bergerheim U, Ekman P, DeMarzo AM, Tibshirani R, Botstein D, Brown PO, Brooks JD, Pollack JR: **Gene expression profiling identifies clinically relevant subtypes of prostate cancer.** *Proc Natl Acad Sci U S A* 2004, **101**:811-816.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

